

InterFaceGAN: Interpreting the Disentangled Face Representation Learned by GANs

Year: 2020 | Citations: 663 | Authors: Yujun Shen, Ceyuan Yang, Xiaoou Tang, Bolei Zhou

Abstract

Although generative adversarial networks (GANs) have made significant progress in face synthesis, there lacks enough understanding of what GANs have learned in the latent representation to map a random code to a photo-realistic image. In this work, we propose a framework called InterFaceGAN to interpret the disentangled face representation learned by the state-of-the-art GAN models and study the properties of the facial semantics encoded in the latent space. We first find that GANs learn various semantics in some linear subspaces of the latent space. After identifying these subspaces, we can realistically manipulate the corresponding facial attributes without retraining the model. We then conduct a detailed study on the correlation between different semantics and manage to better disentangle them via subspace projection, resulting in more precise control of the attribute manipulation. Besides manipulating the gender, age, expression, and presence of eyeglasses, we can even alter the face pose and fix the artifacts accidentally made by GANs. Furthermore, we perform an in-depth face identity analysis and a layer-wise analysis to evaluate the editing results quantitatively. Finally, we apply our approach to real face editing by employing GAN inversion approaches and explicitly training feed-forward models based on the synthetic data established by InterFaceGAN. Extensive experimental results suggest that learning to synthesize faces spontaneously brings a disentangled and controllable face representation.